

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLES

Supplementary Table S1. *rrn5* loci predicted in mitochondrial and plastid genomes and the evidence of *rrn5* gene expression from high-throughput transcriptome data.

Spreadsheets **Mitochondrial-rrn5** and **Plastid-rrn5**

Column ‘Species’: names highlighted in yellow, *rrn5* predicted using a CM developed in this study.

Column ‘Annotation’: +, CM concordantly identified an already annotated *rrn5* locus (up to 2 nt difference of 5’ and 3’ end assignment); +/-, ambiguous identification (i.e. ends shifted by more than 5 nt at either end); -, previously unassigned *rrn5* gene.

Column ‘Genomic position’: By default, coordinates are shown for the hit by the mt-5S model. If the model did not detect the *rrn5* locus above the threshold, mtAT-5S model hit coordinates apply (e.g. oomycetes). Where investigated, experimentally determined termini are given (e.g. malawimonads). For below-threshold hits (*rrn5*-like loci), coordinates were manually curated. If multiple identical *rrn5* loci are present in a genome, coordinates are shown for one arbitrarily chosen locus. +, direct orientation; -, reverse-complement orientation.

Column ‘Neighboring genes’: Upstream and downstream genes are indicated. Arrows pointing to the right or left indicate that neighboring genes are encoded on the same or opposite DNA strand, respectively, as *rrn5*.

Column ‘Expression; SRA dataset’: Transcription of the locus. When SRA datasets were used, the corresponding accession numbers are given. In bold, data generated by us. See Material and Methods for further details.

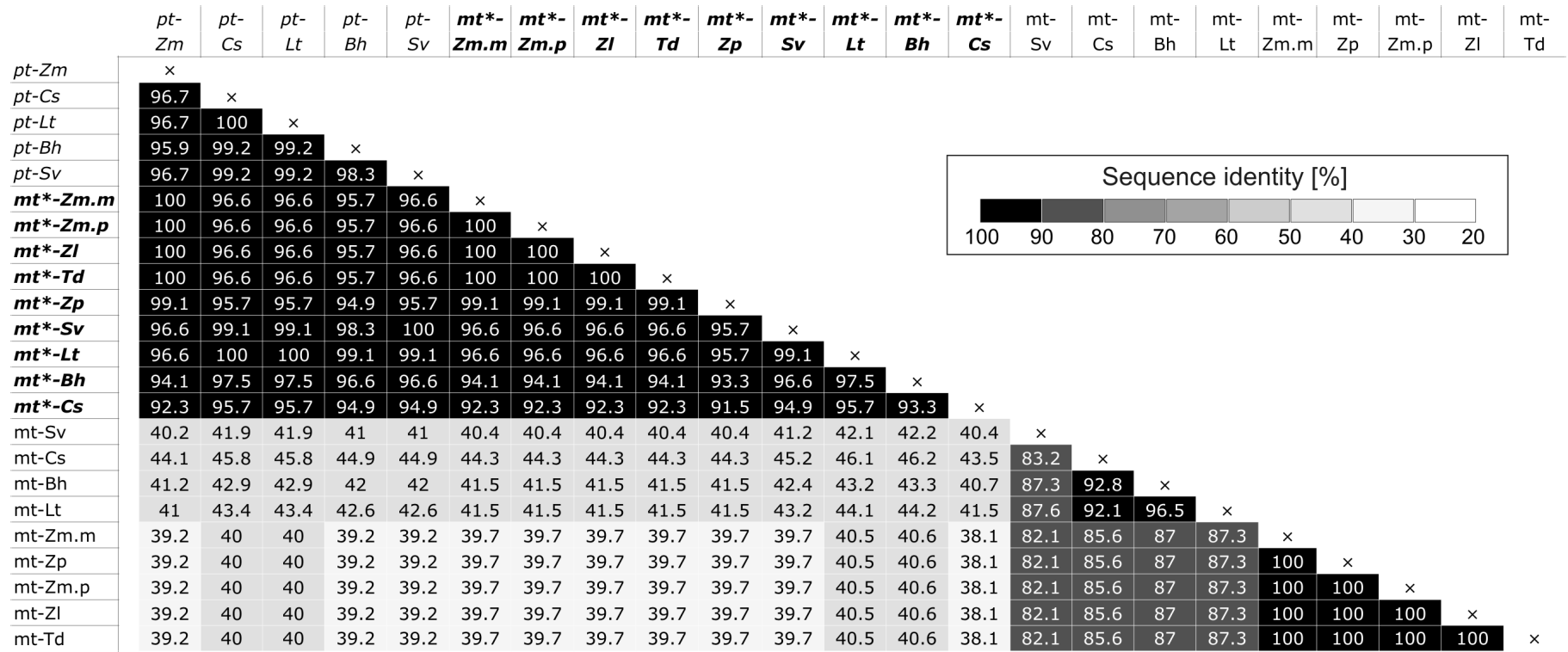
Column ‘Score’ and ‘E-value’: Values, as reported by the model, are shown for the best hit. -, not reported or outside the inclusion values (for details, see Methods). With the mtPerm-model, values are shown only where applicable. n.a., not applicable.

Note that the **Plastid-rrn5** spreadsheet only shows newly identified plastid genome-encoded *rrn5* loci. Also shown are pt-*rrn5* sequences transferred to the mitochondrial genome in nine green plants.

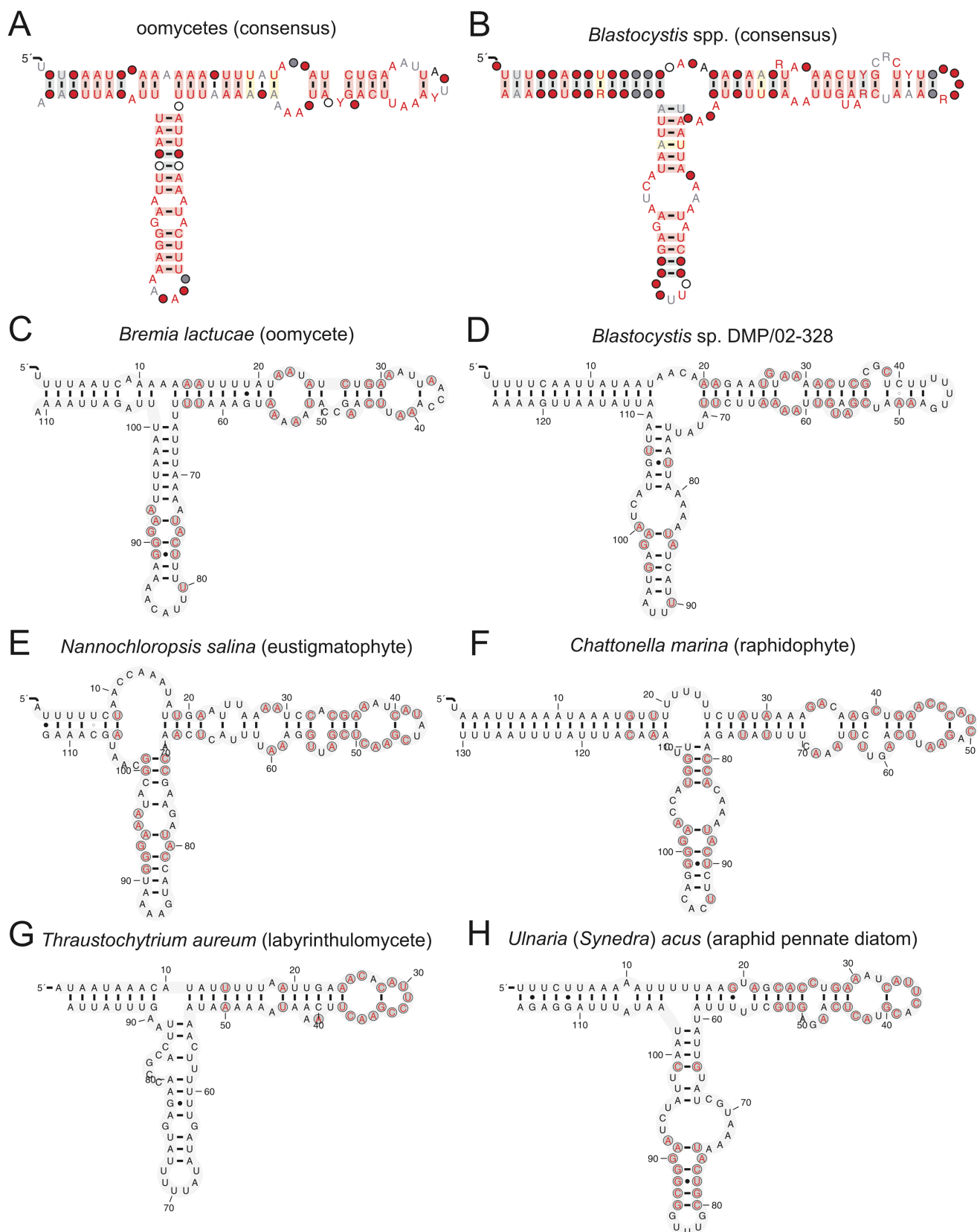
Spreadsheet **rrn5-expression**

Reads were mapped onto mitochondrial or apicoplast genomes and RPKM (reads per kilobase per million of mapped reads) values were calculated to compare the expression of *rrn5* to that of other genes in a given genome. For details, see Methods and spreadsheets **Mitochondrial-rrn5** and **Plastid-rrn5**.

SUPPLEMENTARY FIGURES

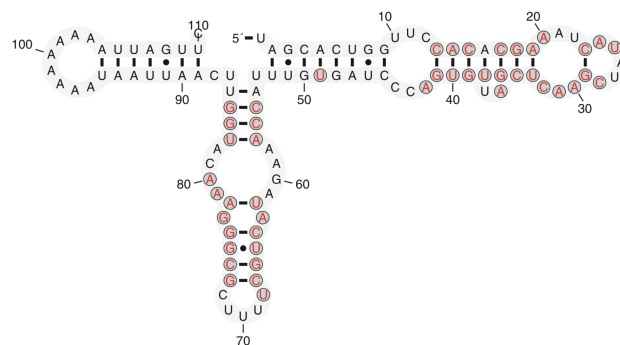
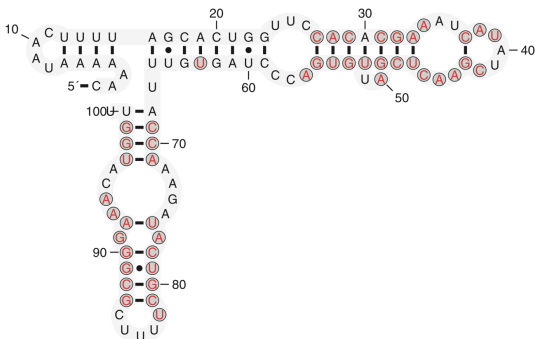


Supplementary Figure S1. Inference of plastid origin of *rrn5* loci identified by the pt-5S model in plant mitochondrial genomes. Plastid *rrn5* sequences (prefix ‘pt-’) were aligned to pt-5S hits in mtDNAs (‘mt*’) and mt-5S hits in mtDNAs (‘mt-’) as described in Methods. The resulting distance matrix is shown as a sequence identity heatmap. Note that a plastid genome sequence is not available for every species with a sequenced mtDNA. Abbreviations of species names: Zm, *Zea mays*; Cs, *Cucumis sativus*; Lt, *Liriodendron tulipifera*; Bh, *Boea hygrometrica*; Sv, *Silene vulgaris*; Zm.m, *Z. mays* subsp. *mays*; Zm.p, *Z. mays* subsp. *parviglumis*; Zl, *Zea luxurians*; Td, *Tripsacum dactyloides*; Zp, *Zea perennis*.

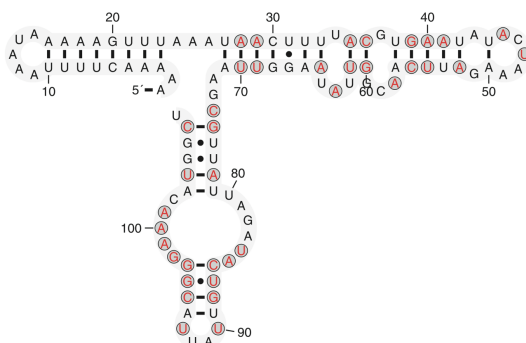


Supplementary Figure S2. (continued)

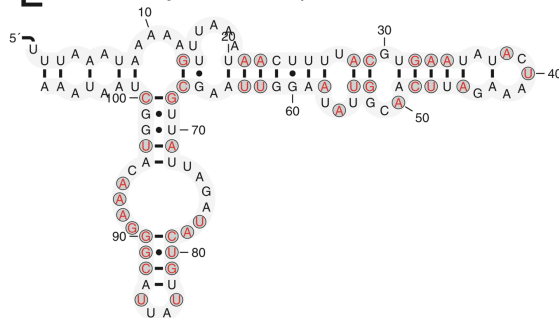
I *Aureococcus anophagefferens* (pelagophyte) | permuted J *A. anophagefferens* | permuted (alternative)



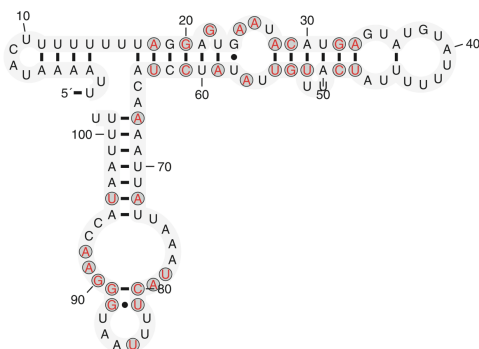
K *Chrysodidymus synuroideus* (synurid) | permuted



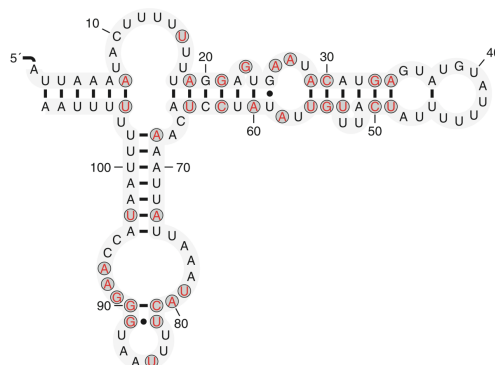
L *C. synuroideus* | conventional



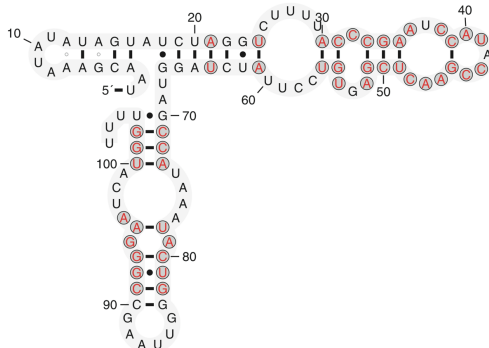
M *Ochromonas danica* (chrysophyte) | permuted



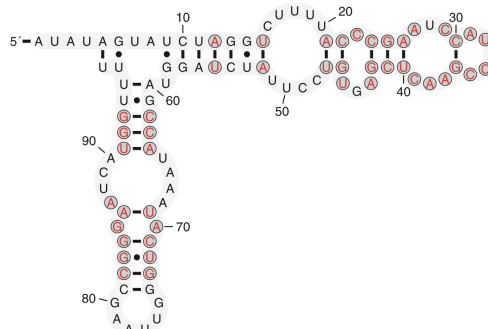
N *O. danica* | conventional



O *Dictyota dichotoma* (phaeophyte) | permuted

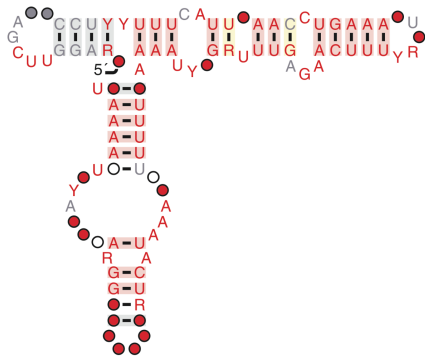


P *D. dichotoma* | conventional

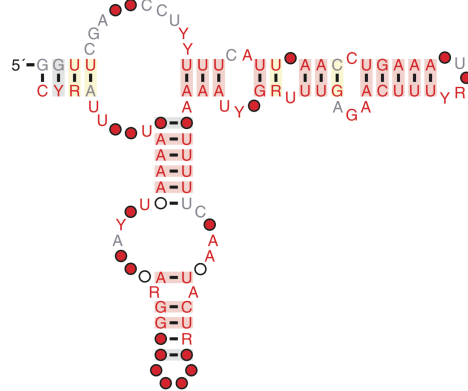


Supplementary Figure S2. (continued)

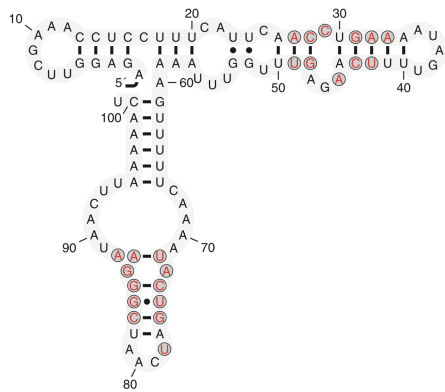
Q raphid pennate diatoms (consensus) | permuted



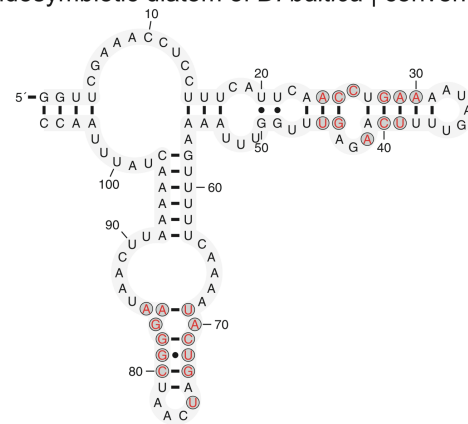
R raphid pennate diatoms (consensus) | conventional



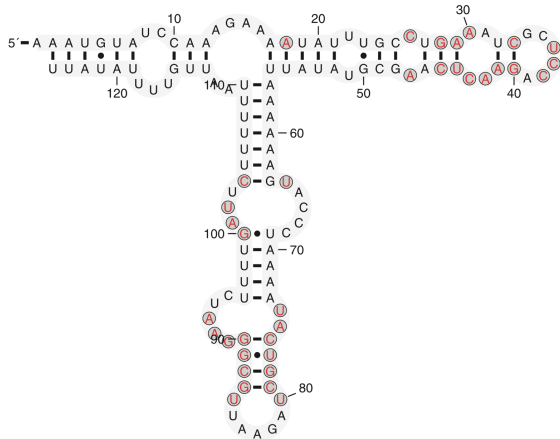
S endosymbiotic diatom of *Durinskia baltica* | permuted



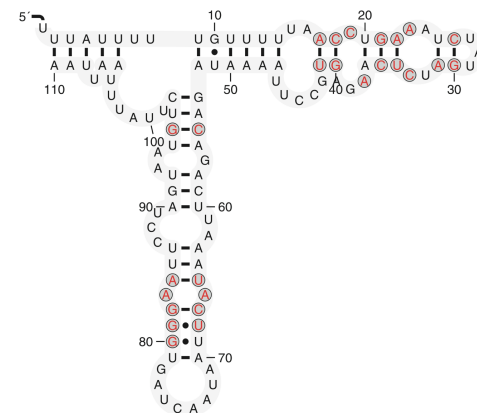
T endosymbiotic diatom of *D. baltica* | conventional



U *Phaeodactylum tricornutum* (raphid pennate diatom)

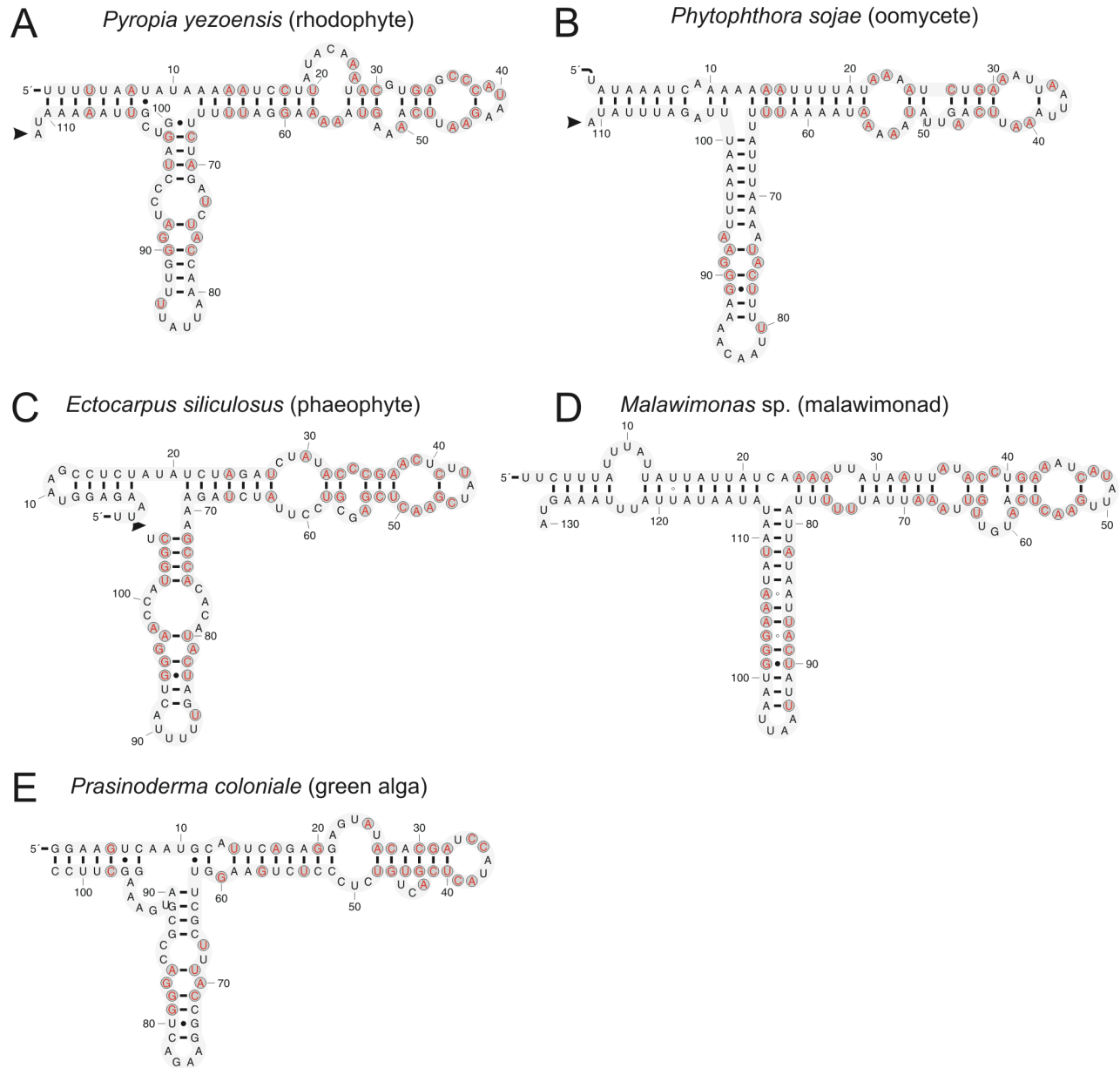


V *Thalassiosira pseudonana* (centric diatom)

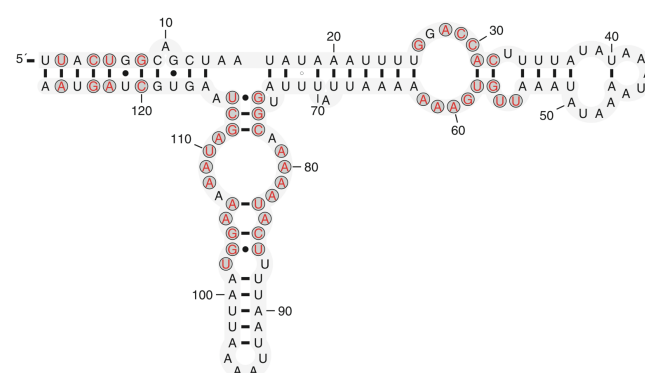
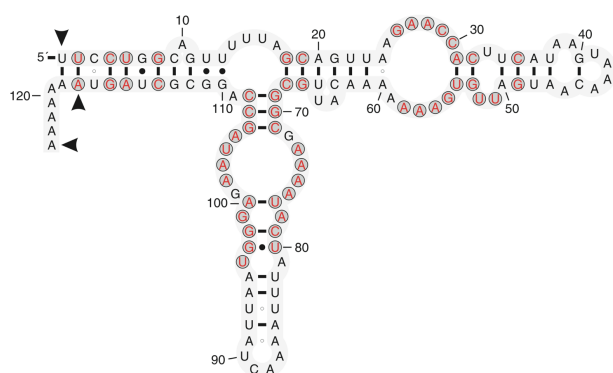
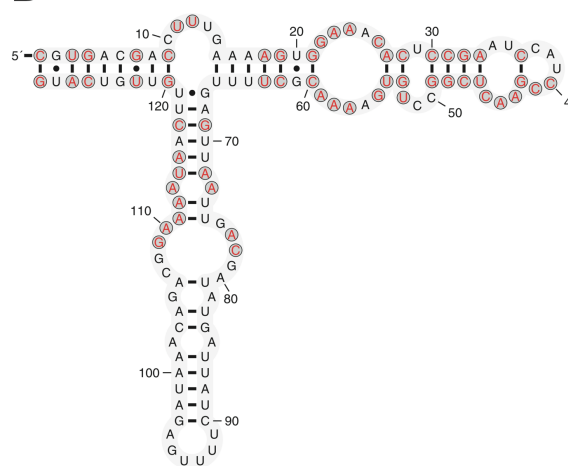
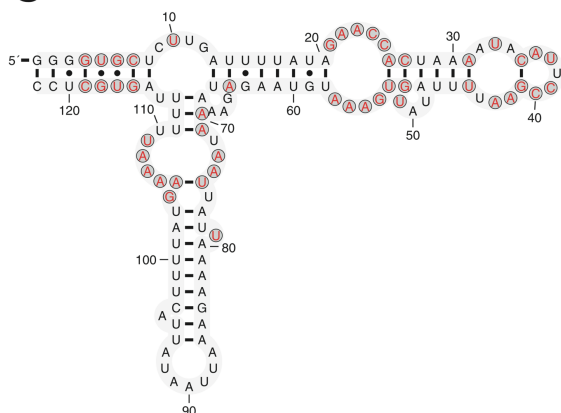
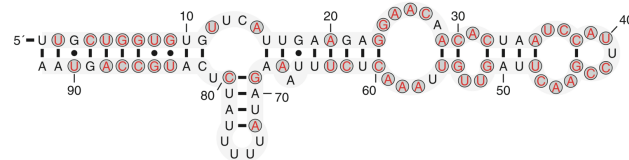
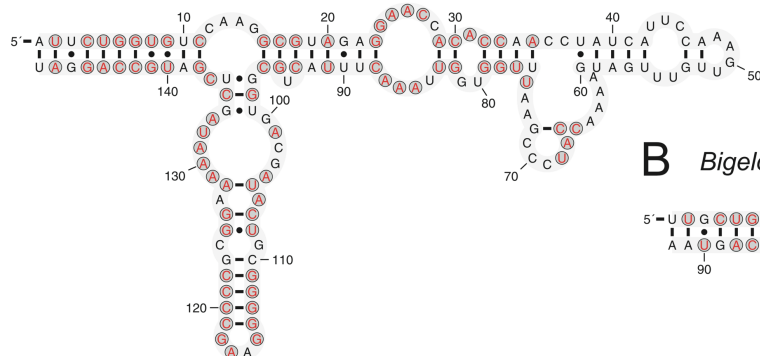


Supplementary Figure S2. Alternative secondary structure models of mt-5S rRNA and 5S-like RNAs from stramenopiles. Circled red nucleotides are conserved relative to the mt-5S model. (A) Oomycetes (consensus; mt-5S rRNA). (B) *Blastocystis* species complex (consensus; 5S-like RNA). (C) *Bremia lactucae* (the highest-scoring oomycete locus). (D) *Blastocystis* sp. DMP/02-328 (the highest-scoring *Blastocystis* locus). (E) *Nannochloropsis salina* (the highest-scoring eustigmatophyte locus). (F) *Chatonella marina* (the highest-scoring raphidophyte locus). (G) *Thraustochytrium aureum*

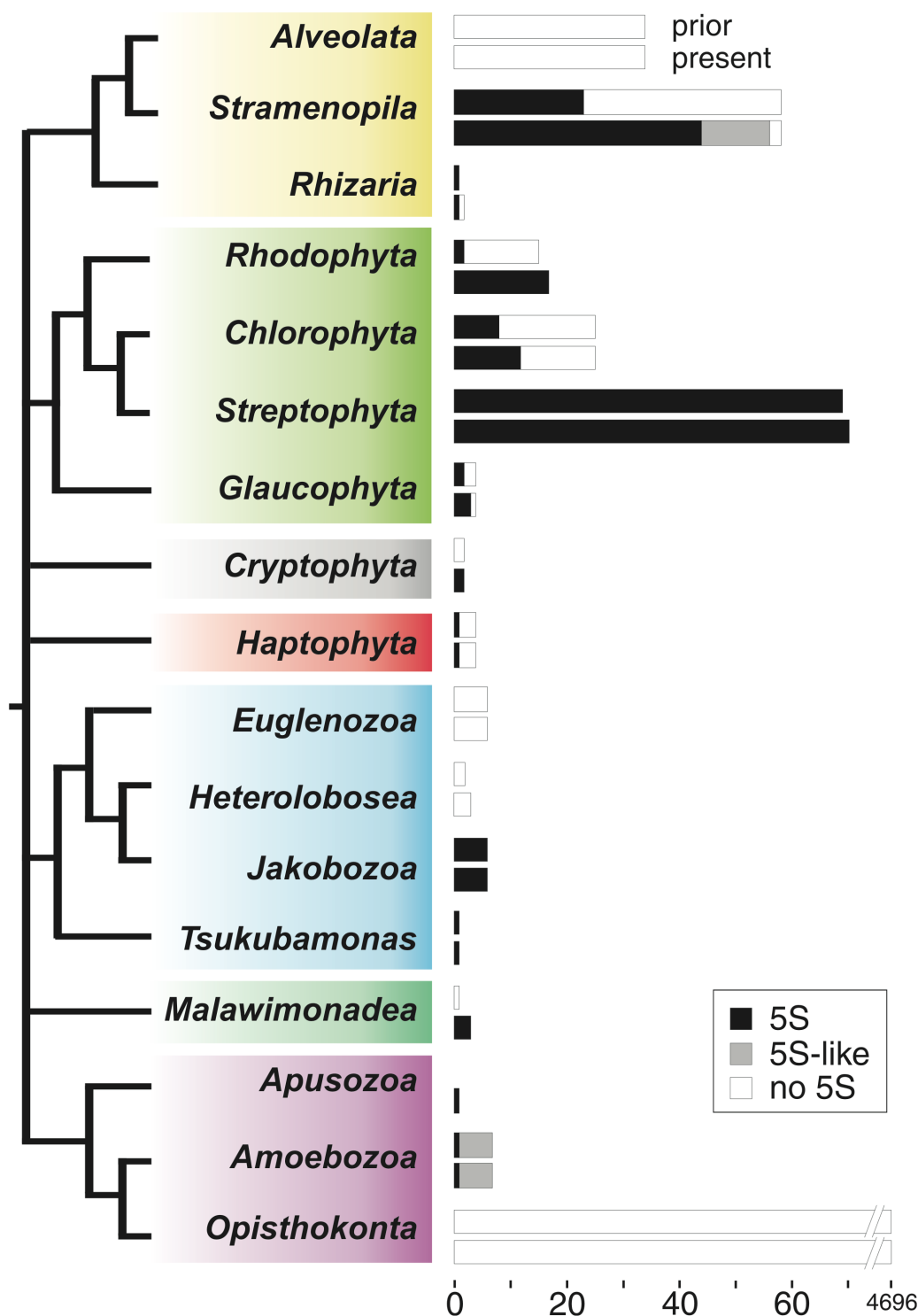
(labyrinthulomycete). **(H)** *Ulnaria (Synedra) acus* (araphid pennate diatom). **(I-J)** *Aureococcus anophageferens* (pelagophyte; 5' permuted structure model as in phaeophytes and an alternative folding with a 3' permuted α domain). **(K-L)** *Chrysodidymus synuroideus* (synurid; 5' permuted model as in phaeophytes and an alternative folding with a conventional α domain). **(M-N)** *Ochromonas danica* (chrysophyte; 5' permuted model as in phaeophytes and an alternative with a conventional α domain). **(O-P)** *Dictyota dichotoma* (phaeophyte; 5' permuted model as in other phaeophytes and an alternative with a conventional α domain). **(Q-R)** raphid pennate diatoms (consensus, without the divergent *P. tricornutum* locus shown in **U**; 5S-like RNA with a 5' permuted vs. reduced conventional α domain). **(S-T)** endosymbiotic diatom of *Durinskia baltica* (the highest-scoring raphid pennate diatom locus; 5' permuted vs. conventional α domain). **(U)** *Phaeodactylum tricornutum* (raphid pennate diatom with an extended conventional α domain and extended γ domain). **(V)** *Thalassiosira pseudonana* (centric diatom with a divergent 5S-like RNA). For **U** and **V**, see [Figure 6](#) and [Supplementary Table S1](#) for evidence of expression.



Supplementary Figure S3. Secondary structure models of mt-5S rRNAs confirmed by transcriptome analyses and models of mt-5S rRNAs with extreme nucleotide composition bias. Black arrowheads indicate positions of the major 3' end seen in RNA-Seq data (**Figure 6**). Circled red nucleotides are conserved relative to the mt-5S model. (**A**) *Pyropia yezoensis* (red alga). (**B**) *Phytophthora sojae* (oomycete). (**C**) *Ectocarpus siliculosus* (brown alga). (**D**) *Malawimonas* sp. (malawimonads; high A+T content). (**E**) *Prasinoderma coloniale* (green alga; high G+C content).



rrn5 among the four gnetophytes). **(B)** *Bigelowiella natans* (chlorarachniophyte). **(C)** *Euglena longa* (euglenid). **(D)** *Chromera velia* (chromerid; the highest-scoring gene copy in its plastid genome is shown). **(E)** *Toxoplasma gondii* (coccidian). Black arrowheads indicate positions of the major ends based on RNA-Seq read mapping (**Figure 6F**). **(F)** *Eimeria tenella* (coccidian). Bottom row, secondary structure skeleton models of the pt-5S rRNA consensus and of **A-F**. Grey shading indicates nucleotide conservation relative to the pt-5S model. For genomic locations, see **Supplementary Table S1**.



Supplementary Figure S5. Biased taxonomic distribution of mtDNA-encoded 5S rRNA across eukaryotes is caused by sampling bias. Length of upper and lower bar represents the number of available mitochondrial genomes prior to and in the present study, respectively. Black, grey, and white filling indicate the proportion of genomes harboring *rrn5*, *rrn5*-like, and no detectable *rrn5* gene, respectively. For a detailed listing of species and corresponding search results, see [Supplementary Table S1](#).